The EMBO Journal vol.7 no.6 pp.1697 - 1704, 1988

Primary structure and processing of lysosomal α-glucosidase; homology with the intestinal sucrase - isomaltase complex

Lies H.Hoefsloot, Marianne Hoogeveen-Westerveld, Marian A.Kroos, Jos van Beeumen[1], Arnold J.J.Reuser and Ben A.Oostra

Department of Cell Biology and Genetics, Erasmus University, 3000 DR Rotterdam, The Netherlands and [1]Laboratory for Microbiology, State University, B-9000 Gent, Belgium

Communicated by A.J. van der Eb

Lysosomal α-glucosidase (acid maltase) is essential for degradation of glycogen in lysosomes. Enzyme deficiency results in glycogenosis type II. The amino acid sequence of the entire enzyme was derived from the nucleotide sequence of cloned cDNA. The cDNA comprises 3636 nt, and hybridizes with a messenger RNA of ~3.6 kb, which is absent in fibroblasts of two patients with glycogenosis type II. The encoded protein has a molecular mass of 104.645 kd and starts with a signal peptide. Sites of proteolytic processing are established by identification of N-terminal amino acid sequences of the 110-kd precursor, and the 76-kd and 70-kd mature forms of the enzyme encoded by the cDNA. Interestingly, both amino-terminal and carboxy-terminal processing occurs. Sites of sugar-chain attachment are proposed. A remarkable homology is observed between this soluble lysosomal α-glucosidase and the membrane-bound intestinal brush border sucrase - isomaltase enzyme complex. It is proposed that these enzymes are derived from the same ancestral gene. Around the putative active site of sucrase and isomaltase, 10 out of 13 amino acids are identical to the corresponding amino acids of lysosomal α-glucosidase. This strongly suggests that the aspartic acid residue at this position is essential for catalytic function of lysosomal α-glucosidase.

Key words: acid α-glucosidase/cloning/glycogenosis II/isomaltase/lysosomal

Introduction

Lysosomal α-glucosidase (acid α-glucosidase, acid maltase EC 3.2.1.20) is a γ-amylase with specificity for glycogen and several natural and synthetic oligoglucosides. The enzyme is most active at low pH (4.0-5.0), and hydrolyzes both α-1,4 and α-1,6 glycosidic linkages (Jeffrey *et al.*, 1970a,b; Palmer, 1971a,b; Rosenfeld, 1975; Koster and Slee, 1977). Deficiency of acid α-glucosidase in mammals and birds (for review, see Walvoort, 1983) leads to accumulation of glycogen in the lysosomes and results in glycogenosis type II (GSD II or Pompe's disease), an autosomal recessive disease (Hers, 1963). The acid α-glucosidase gene has been mapped to chromosome 17, q21-23 (Solomon *et al.*, 1979; Martiniuk *et al.*, 1985).

The clinical presentation of GSD II in humans varies considerably, which has led to classification of subtypes. Infantile GSD II is a rapidly progressive disorder with onset of symptoms shortly after birth. Cardiorespiratory insufficiency is the major cause of death in the first or second year of life. In milder late-onset forms of the disease, the heart is not affected and the main symptom is skeletal muscle weakness (Howell and Williams, 1983). Clinical heterogeneity originates primarily from different mutations of acid α-glucosidase, but secondary genetic and epigenetic factors may be involved (Beratis *et al.*, 1983; Reuser *et al.*, 1985, 1987).

In the course of studying the molecular basis of clinical heterogeneity, information was obtained on the biosynthesis of normal and mutant acid α-glucosidase. *In vitro* translation of acid α-glucosidase mRNA showed an apparent mol. wt of 100 kd for the unglycosylated enzyme precursor. The precursor enters the lumen of the endoplasmic reticulum co-translationally and is glycosylated. This results in a 110-kd molecular species (Van der Horst *et al.*, 1987). The biosynthesis of acid α-glucosidase in cultured fibroblasts proceeds with transfer of the enzyme to the Golgi complex where high-mannose type oligosaccharide side-chains are phosphorylated (Hasilik and Neufeld, 1980a,b; Reuser *et al.*, 1985). Subsequent binding of the precursor to the mannose-6-phosphate receptor ensures further transport to the lysosomes, where mature acid α-glucosidase species of 76 and 70 kd are most abundant. Maturation is a result of proteolytic processing, but the sites of processing and the mechanism are unknown. Acid α-glucosidase species of 105, 100 and 95 kd have been identified as processing intermediates (Reuser *et al.*, 1985; Oude Elferink *et al.*, 1985).

Studies on the biosynthesis of lysosomal α-glucosidase in fibroblasts from patients have led to the conclusion that different mutations are underlying functional enzyme deficiency. Among these are mutations with a specific effect on enzyme synthesis, phosphorylation, proteolytic processing, catalytic activity and intracellular localization of acid α-glucosidase (Reuser *et al.*, 1985, 1987).

Cloning of the cDNA coding for acid α-glucosidase was undertaken to obtain information on the primary structure of the enzyme, to study the maturation process and to investigate in a later stage the exact nature of the mutation in a number of mutant phenotypes. Recently the isolation of a partial cDNA coding for acid α-glucosidase was described (Martiniuk *et al.*, 1986). We now report the full-length coding sequence for acid α-glucosidase. Sites of amino-terminal processing are established, and the attachment site of two sugar chains is predicted. Homology is shown between human lysosomal α-glucosidase, the rabbit intestinal sucrase-isomaltase complex and the partially cloned human isomaltase. Implications for the evolutionary origin of these proteins are discussed.

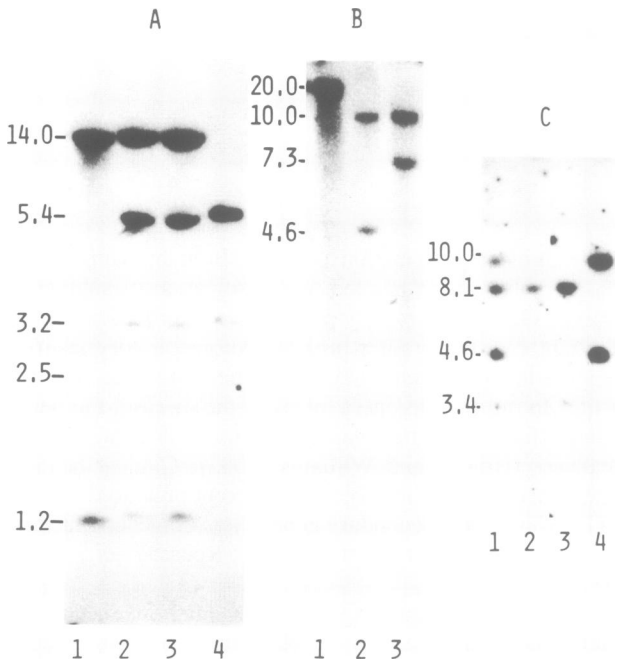

Fig. 1. Southern blot probed with 2.0-kb acid α-glucosidase cDNA. **(A)** *Bam*HI-digested human placenta DNA (**lane 1**), 17CB18A DNA (**lane 2**), 17CB26B DNA (**lane 3**) and Chinese hamster (A23) DNA (**lane 4**). **(B)** Human placenta DNA digested with *Eco*RI (**lane 1**), *Hin*dIII (**lane 2**) and *Bgl*II (**lane 3**). **(C)** *Hin*dIII-digested 60C2HAT DNA (**lane 1**), 60C2BUdR DNA (**lane 2**), mouse LTK$^-$ DNA (**lane 3**) and human placenta DNA (**lane 4**). Fragment lengths are indicated in kilobases.

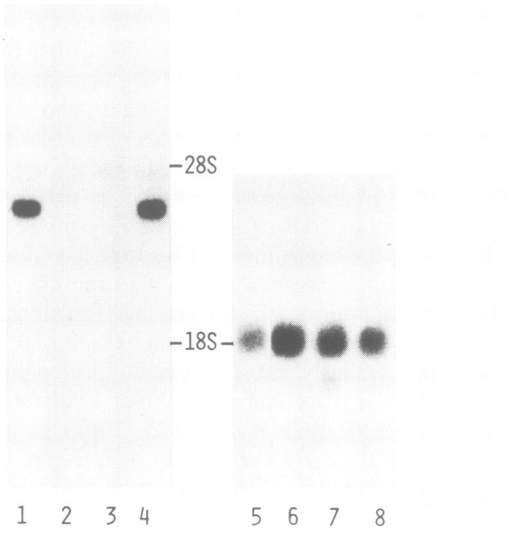

Fig. 2. Northern blot of total RNA extracted from control fibroblasts (**lane 1**), and RNA extracted from fibroblasts of patients with infantile GSD II (**lanes 2–4**). The same blot was hybridized with an actin probe (**lanes 5–8**). The position of the rRNA bands of 28S and 18S is indicated.

Results

Isolation of cDNA clones

A λgt11 human testis cDNA library was screened with a polyclonal antibody preparation against human placenta acid α-glucosidase, and nine clones remained positive after the rescreen. The inserts varied in length from 1.3 to 2 kb. Five of the nine inserts cross-hybridized. The longest of these (2.0 kb) was used as a probe on a Southern blot to test its hybridization with DNA from human chromosome 17 which carries the acid α-glucosidase locus. For this purpose, DNA of hamster–human somatic cell hybrids with no other human chromosomes than 17 was digested with *Bam*HI (Figure 1A). The 2.0-kb insert hybridized with 14- and 1.25-kb *Bam*HI fragments of placenta DNA (lane 1), which fragments were also present in the DNA of the two hybrid cell lines (lanes 2 and 3), but not in DNA of the Chinese hamster cell line A23 (lane 4). Instead, the probe hybridized with 5.4-, 3.2- and 2.5-kb *Bam*HI fragments of Chinese hamster DNA. These same fragments were present in the DNA of the two hybrids, but not in placenta DNA.

Figure 1B shows that the 2.0-kb probe hybridized with a 20-kb fragment of *Eco*RI-digested placenta DNA (lane 1). Hybridization with *Hin*dIII-digested DNA revealed 10- and 4.6-kb fragments (lane 2), and fragments of 10 and 7.3 kb were recognized when placenta DNA was digested with *Bgl*II (lane 3).

For further identification, the 2.0-kb fragment was used as a probe on two different human–mouse hybrid cell lines (Figure 1C). One of these, 60C2HAT, contains only a small fragment of human chromosome 17 and has human acid α-glucosidase activity. The other, 60C2BUdR, is derived from 60C2HAT by selecting against this chromosomal fragment, and has lost activity for human acid α-glucosidase. In DNA of 60C2HAT (lane 1) the probe hybridized with the human 10- and 4.6-kb *Hin*dIII fragments that are also present in placenta DNA (lane 4). However, in DNA extracted from 60C2BUdR (lane 2), no human-derived fragments could be observed to hybridize with the 2.0-kb probe, although the mouse fragments, present in LTK$^-$ (lane 3) and 60C2HAT were detected.

Northern blot analysis of total RNA extracted from control fibroblasts showed a single hybridizing species of 3.6 kb (Figure 2, lane 1), which is long enough to code for a protein of 100 kd. In the same experiment RNA extracted from fibroblasts of three infantile GSD II patients was analyzed. No acid α-glucosidase was synthesized in fibroblasts of two of these patients as judged by immunoblotting (data not shown). In these two cases, no hybridizing mRNA was detected on the Northern blot, not even after a 3-fold increase in exposure (lanes 2 and 3). In cells of the third patient with a normal synthesis of inactive acid α-glucosidase, a normal amount of acid α-glucosidase mRNA was detected (lane 4). Similar amounts of RNA were applied in each lane, as judged from the signal obtained after hybridization of the same blot with an actin probe (lanes 5–8). No gene deletions were discovered by Southern blot analysis of DNA from these patients (data not shown). It is concluded from the analysis of cell hybrids and the Northern blot that the 2.0-kb insert represents part of the cDNA for human acid α-glucosidase.

Nucleotide and deduced amino acid sequences

In order to obtain longer clones, the original clone was used to screen a human placenta λgt11 cDNA library. Several positive clones were obtained and mapped using different restriction enzymes. Five overlapping clones appeared to have the same 3′ end, three of these extended beyond the 5′ end of the 2.0-kb clone. The longest clone was 3.3 kb. One overlapping clone of 2.9 kb missed the 3′ end but extended an extra 300 bp to the 5′ end. Fragments of these

Fig. 3. Nucleotide sequence of the combined cDNA clones coding for acid α-glucosidase. The first in-frame start (ATG) and stop (TAG) codon are underlined, as well as the polyadenylation signal.

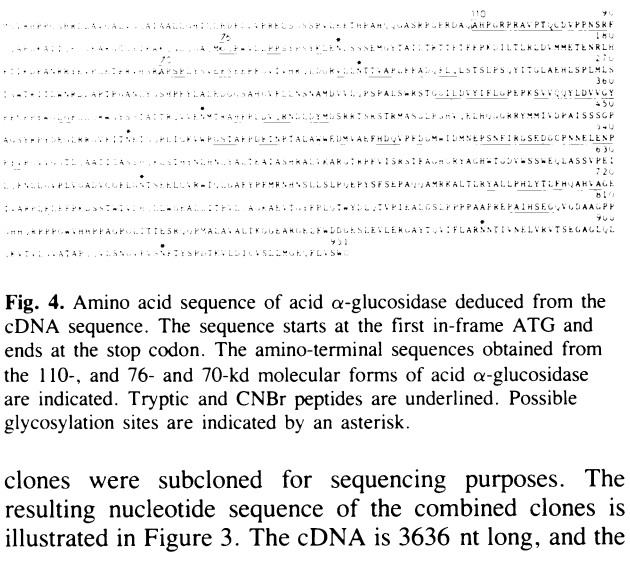

Fig. 4. Amino acid sequence of acid α-glucosidase deduced from the cDNA sequence. The sequence starts at the first in-frame ATG and ends at the stop codon. The amino-terminal sequences obtained from the 110-, and 76- and 70-kd molecular forms of acid α-glucosidase are indicated. Tryptic and CNBr peptides are underlined. Possible glycosylation sites are indicated by an asterisk.

clones were subcloned for sequencing purposes. The resulting nucleotide sequence of the combined clones is illustrated in Figure 3. The cDNA is 3636 nt long, and the longest open reading frame with a length of 2877 nt starts at position 196 and ends with a stop codon at position 3073. In the open reading frame the first ATG is at position 220, therefore the cDNA codes for a polypeptide of 951 amino acid residues. The sequence around this ATG codon matches closely the consensus sequence for translation−initiation sites [(GCC)GCCA/GCCATGG] (Kozak, 1987). A polyadenylation signal (AATAAA) at position 3599 is followed 16 nt later by a stretch of adenosine residues. This indicates that the complete coding sequence and the entire 3′ untranslated region (UTR) was cloned.

The amino acid sequence deduced from the open reading frame from position 220 to position 3073 is shown in Figure 4. Four of the 10 amino acids after the first methionine are basic. These are followed by a stretch of 27 amino acids, 20 of which are hydrophobic. A serine is located in this hydrophobic stretch. These features of the amino acid sequence fulfil the requirements for a signal peptide common to lysosomal and secretory proteins (Watson, 1984).

Sequences of several peptides either obtained by tryptic digestion or CNBr cleavage of human placenta acid α-glucosidase were identified (underlined in Figure 4). The amino-terminal sequence of the 110-kd precursor isolated from human urine and the amino-terminal sequence of the 76- and 70-kd mature forms isolated from placenta were found to start at amino acids 70, 123 and 204 respectively. Seven possible glycosylation sites (Asn-X-Ser/Thr) are indicated by an asterix (Figure 4). The asparagine residues at positions 140 and 233 were not detected in the amino acid sequence analysis. All other asparagine residues in the various peptides gave a clear signal.

Amino acid homology

Comparison of the amino acid sequences of acid α-glucosidase, human isomaltase and rabbit sucrase−isomaltase showed a remarkable homology between these proteins. Sucrase−isomaltase (EC 3.2.1.48−10) is an enzyme complex anchored in the apical membrane of intestinal epithelial cells. The complex is synthesized as a single chain precursor (prosucrase−isomaltase) which is proteolytically cleaved when exposed to pancreatic proteases in the intestinal lumen. The resulting sucrase and isomaltase subunits stay associated with one another in a complex by non-covalent linkages. They have different substrate specificities for sucrose and isomaltose respectively, but share activity for maltose (Brunner *et al.*, 1979; Hauri *et al.*, 1979; Montgomery *et al.*, 1981; Spiess *et al.*, 1982; for review, see Semenza, 1986). The two subunits have 41% of their amino acids identical when they are aligned (Hunziker *et al.*, 1986). In Figure 5 we have aligned the amino acid sequence of acid α-glucosidase with the rabbit isomaltase and sucrase subunit (Hunziker *et al.*, 1986), and the recently published partial sequence of the human isomaltase (Green *et al.*, 1987). Identical amino acids in the sequences are given in boxes; they comprise 26% (246 out of 940). Clusters of identical amino acids occur throughout the sequence. The homology around positions 195, 330, 350, 460, 560 and 720 is most evident (Figure 5). Especially noteworthy are the identical sequences around aspartic acid at position 560, assigned to the active site of sucrase and isomaltase (Quaroni and Semenza, 1976; Hunziker *et al.*, 1986). As much as 10 out of 13 amino acid residues in all four sequences are identical in the region surrounding this aspartic acid.

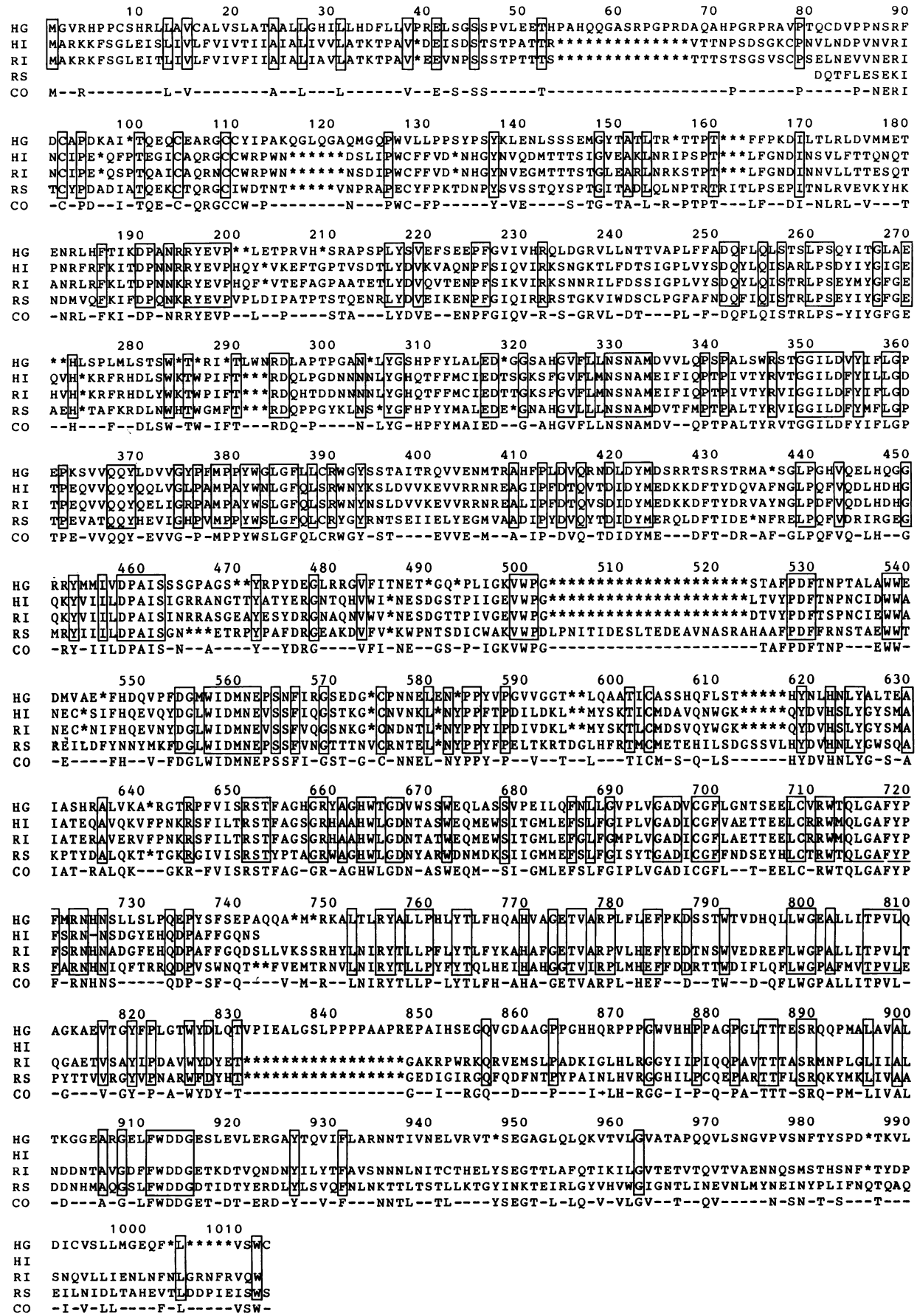

Fig. 5. Amino acid homology of human acid α-glucosidase (HG), human isomaltase (HI), rabbit isomaltase (RI) and rabbit sucrase (RS). Boxes are placed around amino acids identical in all four sequences. Consensus sequence (CO): amino acids identical in two or more sequences are indicated, whereby identical amino acids in human and rabbit isomaltase are counted as 1.

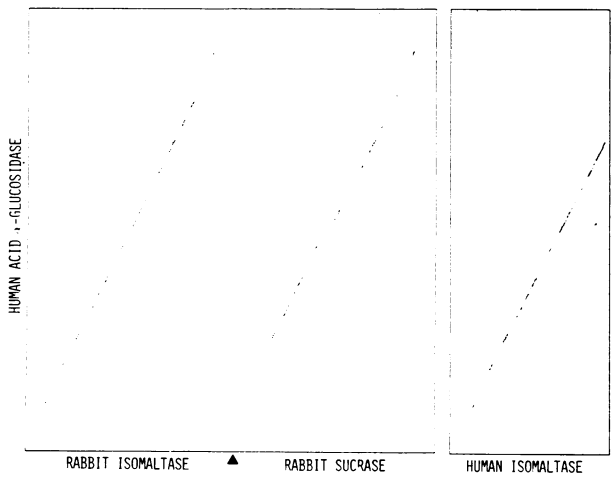

Fig. 6. Comparison of amino acid sequences of acid α-glucosidase, rabbit sucrase–isomaltase and human isomaltase when Dayhoff 'interchanges' are taken into account. Plots were made using Diagon software. [The span was 11 amino acids, the threshold was set at 145. The matrix used was MDM78 from Dayhoff (Staden, 1982).]

Comparison of the amino acid sequence of acid α-glucosidase with each of the other three proteins separately showed that acid α-glucosidase has 41% identical amino acid residues with human isomaltase, 40% with rabbit isomaltase and 36% with rabbit sucrase. When the nucleotides are aligned in the same way, the homology of acid α-glucosidase with human isomaltase, rabbit isomaltase and rabbit sucrase is 44%, 44% and 40% respectively.

The plot in Figure 6, made with Diagon software (Staden, 1982), illustrates the high homology between α-glucosidase and the other proteins when not only identical amino acids but also conservative changes are taken into account.

Discussion

We have isolated cDNA clones encoding human acid α-glucosidase. The combined length of the clones is 3.6 kb, containing the entire coding sequence. The identity of the cDNA was established in several ways. The clone hybridized with DNA sequences from human chromosome 17, particularly with those derived from a small and cytogenetically undetectable chromosomal fragment carrying the acid α-glucosidase gene. Furthermore, the clone hybridized with a messenger of 3.6 kb, which was absent in cells of two patients with infantile GSD II. The unambiguous proof was the identification of the amino-terminal sequences of the 110-kd precursor, the 76- and 70-kd mature forms of acid α-glucosidase, and several tryptic and CNBr peptides in the amino acid sequence deduced from the cDNA clone. These sequences are all encoded in the longest open reading frame, in which only one methionine precedes the amino terminus of the 110-kd precursor. That translation starts, indeed, at this methionine is further indicated by the fact that the ATG codon is flanked by the consensus sequence for translation–initiation sites, and is followed by a putative signal sequence. At the 3′ end of the clone a polyadenylation signal is followed by a stretch of adenosine residues. This implies that the complete coding sequence and the 3′ UTR of acid α-glucosidase was cloned. Since the combined length of the cDNA is similar to the estimated length of the mRNA, we might have cloned the entire 5′ UTR as well, although it is possible that a few nucleotides are missing at the utmost 5′ end.

The cDNA clone recognized restriction fragments in human placenta DNA similar in size to the fragments detected by Martiniuk *et al.* (1986) with their partial acid α-glucosidase cDNA. A 20-kb human *Eco*RI fragment was also observed to hybridize with an acid α-glucosidase cDNA cloned from monkey testis (Konings *et al.*, 1984). However, we have found no cross-hybridization of this previously cloned monkey cDNA and the presently described human cDNA, casting doubt on the identity of the monkey cDNA.

Analysis of DNA and RNA extracted from fibroblasts of patients showed deficiency of mRNA in some cases, but did not reveal gene deletions.

The human cDNA encodes a protein with a mol. wt of 104.645 kd, similar in size to the 100-kd unglycosylated precursor of acid α-glucosidase which is produced by *in vitro* translation (Van der Horst *et al.*, 1987).

Some interesting features concerning the post-translational processing of acid α-glucosidase emerge when the spacing of the sequences coding for the precursor and mature enzyme is considered. The 110-kd amino acid sequence starts 69 amino acids after the first methionine. Since signal peptides are in general not much longer than 30 amino acids (Watson, 1984), the 110-kd precursor isolated from human urine may not be the first *in vivo* precursor. It is likely that some proteolytic processing has occurred after removal of the signal peptide. Indeed, as described by Oude Elferink *et al.* (1984), the precursor present in human urine is slightly different from the precursor in fibroblasts. The precursor found in fibroblasts might well be the first glycosylated precursor, missing only the signal peptide.

The distance between the amino termini of the 110-kd precursor and the 76-kd mature form of acid α-glucosidase is only 53 amino acids (5.5 kd). This implies that the bulk of proteolytic processing necessary to reduce 110 kd to 76 kd occurs at the carboxy-terminal end. The 110-kd precursor starting at position 70 in the amino acid sequence has a calculated mol. wt of 97 kd, when not glycosylated. The 76-kd mature form is glycosylated and contains probably four sugar chains (Belenky *et al.*, 1979; Mutsaers *et al.*, 1987), which account for ~5 kd. This implies that a fragment of ~20 kd (97 − 71 − 5.5) is spliced off at the carboxy-terminal end. Other lysosomal enzymes known to undergo carboxy-terminal processing are cathepsin D and β-glucuronidase (Erickson *et al.*, 1984). However, their carboxy-terminal processing is limited to only 1 kd for cathepsin D and ~3 kd for β-glucuronidase.

The difference between the starting points of the 76- and 70-kd mature species is 78 amino acid residues, or 8.0 kd. This suggests that processing of the 76-kd mature enzyme occurs mainly at the amino-terminal end. Carboxy-terminal sequencing is required to determine the exact end of the mature forms of acid α-glucosidase.

The predicted amino acid sequence between the 76-kd amino terminus and the carboxy-terminal end of the protein contains seven possible sites of N-linked glycosylation. But only four carbohydrate chains are present in the 76- and 70-kd enzymes. It is likely that the asparagine residues at positions 140 and 233 are glycosylated since they were not detected by amino acid sequencing carried out twice. By contrast, all other asparagine residues present in the peptides were detected by amino acid sequencing. The site at position

140 is lost by amino-terminal processing of the 76-kd to the 70-kd form. The two potential glycosylation sites at positions 881 and 924 are lost during carboxy-terminal processing from the 110-kd to the 76-kd form. This implies that two of the three remaining glycosylation sites at positions 390, 469 and 651 are used.

The amino acid sequence of human acid α-glucosidase showed a remarkable homology with the amino acid sequences of rabbit sucrase–isomaltase and human isomaltase. A comparison of the acid α-glucosidase sequence with the protein sequence data library of the National Biochemical Research Foundation (NBRF) revealed no significant homologies with other proteins. The degree of homology between acid α-glucosidase, sucrase and isomaltase is surprisingly high considering the different characteristics of these proteins, although some homology could have been anticipated because of their similar substrate specificities. Acid α-glucosidase is a soluble lysosomal enzyme with optimal activity at a pH between 4 and 5 (Jeffrey *et al.*, 1970a,b; Palmer, 1971a,b; Rosenfeld, 1975; Koster and Slee, 1977), whereas sucrase–isomaltase is anchored in the apical membrane of intestinal epithelial cells and has optimal activity at neutral pH (for review, see Semenza, 1986). Moreover, glycogen is the natural substrate of the lysosomal enzyme, while sucrase–isomaltase does not act on this polysaccharide.

As proposed by Hunziker *et al.* (1986), the isomaltase–sucrase single chain precursor (prosucrase–isomaltase) has arisen by duplication of an ancestral isomaltase gene. Subsequently, the two parts of the new protein evolved into one part still acting on isomaltose, whereas the other acquired activity towards sucrose. The duplication supposedly took place before separation of mammals and reptiles, i.e. more than 3×10^8 years ago. The fact that the partial sequence of human isomaltase has a much higher homology with the rabbit counterpart (85% at the DNA level and 82% at the amino acids level) (Green *et al.*, 1987) supports this hypothesis, since the separation between man and rodents took place only 8×10^7 years ago. Since lysosomal α-glucosidase has an equal degree of homology with isomaltase and sucrase, and acid α-glucosidase is not the product of a duplicated gene, we suggest that acid α-glucosidase evolved from the same ancestral gene before or at the time of the gene duplication.

A partial amino acid sequence of the ancestral protein was postulated by choosing amino acids at each position that occur most frequently in the four aligned proteins (Figure 5). Identical amino acids occurring in the human and rabbit isomaltase sequences were counted as one. This consensus sequence comprises 63% (595 out of 951) of the shared amino acid positions.

The homology between acid α-glucosidase, sucrase and isomaltase allows us to draw some conclusions about the functionally important domains. For instance, the stretch of six identical amino acid residues at positions 557–562 includes the aspartic acid residue that is a constituent of the active site of sucrase and isomaltase (Quaroni and Semenza, 1976; Hunziker *et al.*, 1986). This strongly suggests that these sequences are also part of the active site of acid α-glucosidase. Stretches of identical amino acids occurring at various other positions may be of additional importance for recognition and/or binding of substrate.

Acid α-glucosidase and prosucrase–isomaltase differ most at the amino-terminal and carboxy-terminal ends. Different amino acids are present in the signal peptides of both proteins, and, indeed, they channel the enzymes to a different cellular location. Furthermore, the signal peptide of prosucrase–isomaltase is not cleaved off and serves as a membrane anchor (Hunziker *et al.*, 1986), whereas acid α-glucosidase, like other lysosomal enzymes, loses its signal peptide. The serine and threonine-rich segment of prosucrase–isomaltase, which forms a connecting stalk between the transmembrane and other domains of sucrase–isomaltase, is absent in acid α-glucosidase. Other non-homologous regions at the amino-terminal and carboxy-terminal ends may have a function for enzyme-specific, post-translational processing or may determine the partially different substrate specificities. Computer-assisted comparison of the sequences may reveal further interesting similarities and dissimilarities. Analyses of mutations in acid α-glucosidase in various cases of GSD II will give additional information on which sequences and domains are important for post-translational processing and catalytic function.

Materials and methods

Isolation of cDNA clones

The following cDNA libraries were used. A human testis λgt11 cDNA library was purchased from Clontech Inc. (Palo Alto, CA) with 1×10^6 independent clones and insert lengths between 0.7 and 3.3 kb. A human placenta λgt11 cDNA library (VII-75-1, placenta A1) was a generous gift of Dr J.E.Sadler (Washington University School of Medicine, St Louis, MO). The library consisted of 4×10^6 independent clones with insert lengths varying between 1.0 and 3 kb (Ye *et al.*, 1987).

Screening of libraries

Libraries were screened with antibodies essentially as described by Huynh *et al.* (1985). A polyclonal antiserum raised in rabbit against purified human placenta acid α-glucosidase was used. The IgG fraction was isolated, and antibodies against *Escherichia coli* and phage antigens were absorbed by incubating the antiserum with nitrocellulose filters soaked in lysates of Y1089 containing lysogenic λgt11. Screening of libraries with nucleic acid probes was as described by Benton and Davis (1977). A partial restriction map of the isolated clones is given in Figure 7.

Isolation and analysis of cellular DNA and RNA

Standard procedures were used for extraction of DNA and Southern blot analysis (Maniatis *et al.*, 1982). Cellular RNA was isolated using the lithium chloride method (Auffray and Rougeon, 1980). Total RNA was fractionated on a 1% agarose gel in the presence of formaldehyde and blotted onto nitrocellulose filters. Filters were hybridized in the standard hybridization buffer (Maniatis *et al.*, 1982). After 16 h the filters were washed till a stringency of 0.3 × SSC, and an autoradiograph was made. DNA probes were radioactively labelled as described by Feinberg and Vogelstein (1983).

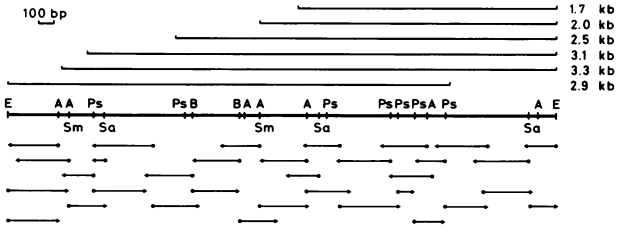

Fig. 7. Cloning and sequencing strategy for acid α-glucosidase. Six isolated clones are indicated. Bottom line: partial restriction map of the combined clones. The 1.7- and 2.0-kb clones were isolated from a human testis library using antibody screening, and the 2.0-kb clone was used to isolate the 2.5-, 2.9-, 3.1- and 3.3-kb clones from a human placenta library. Arrows indicate the sequence strategy followed. Ps, *Pst*I; A, *Ava*I; E, *Eco*RI; Sa, *Sac*I; Sm, *Sma*I.

Sequence analyses of DNA, protein and peptides

Sequencing of cDNA clones was done by subcloning appropriate fragments in pTZ18 or pTZ19 (Pharmacia), using the dideoxy chain termination method with [^{35}S]dATP as labelling component (Biggin *et al.*, 1983). The strategy used for cDNA sequencing is outlined in Figure 7. When no appropriate fragments could be obtained, an oligonucleotide complementary to the cDNA was used as a primer. Oligonucleotides were 22 nt long, and were synthesized on an Applied Biosystems 381A DNA synthesizer. Purification was according to the manufacturer's protocol.

Precursor acid α-glucosidase (110 kd) was purified from human urine as described by Oude Elferink *et al.* (1984) and analyzed for purity by PAGE in the presence of SDS according to Laemmli (Laemmli, 1970). The 76- and 70-kd mature forms of acid α-glucosidase were purified from human placenta as described previously (Reuser *et al.*, 1985). These two species (3 mg) were electrophoretically separated using SDS–PAGE (10% gel). The SDS was purchased from Serva (no. 20760) and recrystallized as described (Hunkapillar *et al.*, 1983). Thioglycolate (0.1 mM) was added to the cathodal buffer compartment. The gel was incubated in 4 M sodium acetate to visualize the protein. Protein bands were excised and sliced into pieces from which the protein was eluted electrophoretically using a Schleicher and Schuell 'BIOTRAP'. The buffer used was 10 mM ammonium bicarbonate (pH 8.2) with addition of 0.01% SDS and 0.05% β-mercaptoethanol. Electroelution was performed for 5 h at room temperature followed by 1 h in the same buffer without SDS. Electroeluted protein was analyzed by SDS–PAGE and used for determination of amino-terminal sequences. Amino-terminal sequences of the 76- and 70-kd species were also obtained after separation on SDS–PAGE followed by electrophoretic transfer to polybrene-coated glass-fibre sheets as described (Vandekerckhove *et al.*, 1985).

Tryptic peptides were obtained from 2 mg of acid α-glucosidase purified from human placenta, containing equal amounts of 76- and 70-kd protein. After reduction and alkylation, the protein was digested with trypsin (Cooper Biomedical, no. 3740) in 0.5 M Tris–HCl (pH 8.5), at 37°C for 48 h using an enzyme/protein ratio of 3:100 (w/w) (Yuan *et al.*, 1982). The resulting peptides were lyophilized and dissolved in acetonitrile (16%), and water (84%) containing 0.1% trifluoroacetic acid, and separated on a 25-cm C8 reversed-phase HPLC column (RP8 Lichrospher, Merck) using a gradient of 0–33% acetonitrile in water with 0.1% trifluoroacetic acid. Peaks were collected and rerun on a 25-cm μBondapak column (Waters Associates) using the same solvent system. Seemingly homogeneous fractions were collected and used for amino acid sequence analysis.

CNBr cleavage of placental acid α-glucosidase was carried out as described by O'Dowd *et al.* (1985). The resulting fragments were separated by SDS–PAGE (13% gel), and electrophoretically transferred to polybrene-coated glass-fibre sheets for sequencing purpose. Amino acid sequence analysis of proteins and peptides was performed on an Applied Biosystems sequencer (477A) on line with the PTH (120A) analyzer.

Hybrid cell lines

Human–hamster somatic cell hybrids were obtained by fusing thymidine kinase deficient (TK^-) Chinese hamster fibroblasts (cell line A23) with human leukocytes. Hybrid clones (17CB18A and 17CB26B) containing human chromosome 17 were selected in HAT medium (Littlefield, 1964) and investigated cytogenetically. The two independent clones containing no other human chromosomes than chromosome 17 were kindly provided by Dr A.Geurts van Kessel, Department of Cell Biology and Genetics, Erasmus University, Rotterdam.

A clone (60C2HAT) of the mouse LTK^- cell line containing a cytogenetically undetectable fragment of human chromosome 17 including the loci for TK, galactokinase (GALK) and acid α-glucosidase (GAA) was obtained by (HeLa) chromosome-mediated gene transfer followed by HAT selection. The clone expressed human acid α-glucosidase activity. Through growth of 60C2HAT in medium containing BUdR a new clone was obtained that no longer contained human chromosome sequences as judged by the loss of activity for acid α-glucosidase and TK (60C2BUdR) (De Jonge *et al.*, 1985).

Acknowledgements

The authors wish to thank Professor Dr Hans Galjaard for his continuous support. Many colleagues at the Department of Cell Biology and Genetics have given valuable advice at different stages. Dr André Hoogeveen is acknowledged for helping with the HPLC separation of tryptic peptides, Sjozèf van Baal, Marcel van Duin and Dr Colin Wynn for computer analyses of sequencing data, and Joop Fengler and Pim Visser for preparing the figures. We thank Dr Reinoud Amons and Jack de Graaf from the SON facility for protein sequencing at the Laboratory for Medical Biochemistry in Leiden for their excellent work, and Dr Dallas Swallow for her splendid suggestion to search for homology between acid α-glucosidase and sucrase–isomaltase. This work was financed in part by the Netherlands organization for scientific research (NWO) under the auspices of the Netherlands Foundation for Medical and Health Research (MEDIGON), and by the Belgium State Prime Minister's Office Science Policy Programming (Belgium National Incentive Program on fundamental research in Life Sciences).

References

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.
Belenky,D.M., Mikhajlov,V.K. and Rosenfeld,E.L. (1979) *Clin. Chim. Acta*, **93**, 365–370.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Beratis,N.G., LaBadie,G.U. and Hirschhorn,K. (1983) *Curr. Top. Biol. Med. Res.*, **11**, 25–36.
Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
Brunner,J., Hauser,H., Braun,H., Wilson,K.J., Wacker,H., O'Neill,B. and Semenza,G. (1979) *J. Biol. Chem.*, **254**, 1821–1828.
De Jonge,A.J.R., De Smit,S., Kroos,M.A. and Reuser,A.J.J. (1985) *Hum. Genet.*, **69**, 32–38.
Erickson,A.H., Conner,G.E. and Blobel,G. (1984) In Barranger,J.A. and Brady,R.O. (eds), *Molecular Basis of Lysosomal Storage Disorders*. Academic Press, New York, pp. 235–249.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Green,F., Edwards,Y., Hauri,H.-P., Povey,S., Ho,M.W., Pinto,M. and Swallow,D. (1987) *Gene*, **57**, 101–110.
Hasilik,A. and Neufeld,E.F. (1980a) *J. Biol. Chem.*, **255**, 4937–4945.
Hasilik,A. and Neufeld,E.F. (1980b) *J. Biol. Chem.*, **255**, 4946–4950.
Hauri,H.-P., Quaroni,A. and Isselbacher,K.J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5183–5186.
Hers,H.G. (1963) *Biochem. J.*, **86**, 11–16.
Howell,R.R. and Williams,J.C. (1983) In Stanbury,J.O., Wyngaarden,J.B. and Frederickson,D.S. (eds), *The Metabolic Basis of Inherited Disease*. McGraw-Hill, New York, pp. 141–166.
Hunkapillar,M.W., Lujan,E., Ostander,F. and Hood,L.E. (1983) *Methods Enzymol.*, **91**, 227–236.
Hunziker,W., Spiess,M., Semenza,G. and Lodish,H.F. (1986) *Cell*, **46**, 227–234.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, Vol. I, pp. 49–78.
Jeffrey,P.L., Brown,D.H. and Brown,B. (1970a) *Biochemistry*, **9**, 1403–1415.
Jeffrey,P.L., Brown,D.H. and Brown,B. (1970b) *Biochemistry*, **9**, 1416–1423.
Konings,A., Hupkes,P., Versteeg,R., Grosveld,G., Reuser,A. and Galjaard,H. (1984) *Biochim. Biophys. Acta*, **119**, 252–258.
Koster,J.F. and Slee,R.G. (1977) *Biochim. Biophys. Acta*, **482**, 89–97.
Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8148.
Laemmli,U.K. (1979) *Nature*, **227**, 680–685.
Littlefield,J.W. (1964) *Science*, **145**, 709–710.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Martiniuk,F., Ellenbogen,A., Hirschhorn,K. and Hirschhorn,R. (1985) *Ann. Hum. Genet.*, **48**, 49–56.
Martiniuk,F., Mehler,M., Pellicer,A., Tzall,S., LaBadie,G., Hobart,C., Ellenbogen,A. and Hirschhorn,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9641–9644.
Montgomery,R.K., Sybicki,M.A., Forcier,A.G. and Grand,R.J. (1981) *Biochim. Biophys. Acta*, **661**, 346–349.
Mutsaers,J.H.G.M., Van Halbeek,H., Vliegenthart,J.F.G., Tager,J.M., Reuser,A.J.J., Kroos,M. and Galjaard,H. (1987) *Biochim. Biophys. Acta*, **911**, 244–251.
O'Dowd,B.F., Quan,F., Willard,H.F., Lamhonwah,A.-M., Korneluk,R.G., Lowden,J.A., Gravel,R.A. and Mahuran,D.J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1184–1188.
Oude Elferink,R.P.J., Brouwer,E.M., Surya,I., Strijland,A., Kroos,M., Reuser,A.J.J. and Tager,J.M. (1984) *Eur. J. Biochem.*, **139**, 489–495.
Oude Elferink,R.P.J., Van Doorn-Van Wakeren,J., Strijland,A., Reuser,A.J.J. and Tager,J.M. (1985) *Eur. J. Biochem.*, **153**, 55–63.
Palmer,T.N. (1971a) *Biochem. J.*, **124**, 701–711.

Palmer,T.N. (1971b) *Biochem. J.*, **124**, 713–724.
Quaroni,A. and Semenza,G. (1976) *J. Biol. Chem.*, **251**, 3250–3253.
Reuser,A.J.J., Kroos,M., Oude Elferink,R.P.J. and Tager,J.M. (1985) *J. Biol. Chem.*, **260**, 8336–8341.
Reuser,A.J.J., Kroos,M., Willemsen,R., Swallow,D., Tager,J.M. and Galjaard,H. (1987) *J. Clin. Invest.*, **79**, 1689–1699.
Rosenfeld,E.L. (1975) *Pathol. Biol.*, **23**, 71–84.
Semenza,G. (1986) *Annu. Rev. Cell Biol.*, **2**, 255–313.
Solomon,E., Swallow,D., Burgess,S. and Evans,L. (1979) *Ann. Hum. Genet.*, **42**, 273–281.
Spiess,M., Brunner,J. and Semenza,G. (1982) *J. Biol. Chem.*, **257**, 2370–2377.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951–2961.
Vandekerckhove,J., Bauw,G., Puype,M., Van Damme,J. and Van Montagu,M. (1985) *Eur. J. Biochem.*, **152**, 9–19.
Van der Horst,G.T.J., Hoefsloot,E.H., Kroos,M.A. and Reuser,A.J.J. (1987) *Biochem. Biophys. Acta*, **910**, 123–129.
Walvoort,H.C. (1983) *J. Inher. Metab. Dis.*, **6**, 3–16.
Watson,M.E.E. (1984) *Nucleic Acids Res.*, **12**, 5145–5164.
Ye,R.D., Wun,T.-C. and Sadler,J.E. (1987) *J. Biol. Chem.*, **262**, 3718–3725.
Yuan,P.-M., Pande,H., Clark,B.R. and Shively,J.E. (1982) *Anal. Biochem.*, **120**, 289–301.

Received on February 15, 1988; revised on March 18, 1988

Note added in proof

These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number Y00839.